

Formation and dissociation of protonated cytosine–cytosine base pairs in i-motifs by *ab initio* quantum chemical calculations*

Zhang Xiao-Hu(张小虎)^{a)}, Li Ming(黎明)^{b)},
Wang Yan-Ting(王延颀)^{a)†}, and Ouyang Zhong-Can(欧阳钟灿)^{a)}

^{a)} State Key Laboratory of Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China

^{b)} Department of Physical Science, University of Chinese Academy of Sciences, Beijing 100190, China

(Received 22 September 2013; revised manuscript received 17 October 2013; published online 20 December 2013)

Formation and dissociation mechanisms of $C-C^+$ base pairs in acidic and alkaline environments are investigated, employing *ab initio* quantum chemical calculations. Our calculations suggest that, in an acidic environment, a cytosine monomer is first protonated and then dimerized with an unprotonated cytosine monomer to form a $C-C^+$ base pair; in an alkaline environment, a protonated cytosine dimer is first unprotonated and then dissociated into two cytosine monomers. In addition, the force for detaching a $C-C^+$ base pair was found to be inversely proportional to the distance between the two cytosine monomers. These results provide a microscopic mechanism to qualitatively explain the experimentally observed reversible formation and dissociation of i-motifs.

Keywords: *ab initio* quantum chemical calculation, i-motif, protonated cytosine–cytosine base pair

PACS: 07.05.Tp, 87.15.A–, 87.15.–v

DOI: 10.1088/1674-1056/23/2/020702

1. Introduction

I-motif is a cytosine-rich DNA oligomer with a tetrameric structure formed in an acidic environment. It is a four-stranded complex consisting of two base-paired parallel-stranded duplexes, each of which is held together by protonated cytosine–cytosine ($C-C^+$) base pairs. In an acidic environment, i-motif, a cytosine-rich DNA oligomer, always adopts a four-stranded structure with two associated and fully intercalated parallel-stranded duplexes, as has been shown by NMR,^[1–3] X-ray diffraction analysis^[4–7] and circular dichroism studies.^[8] The building blocks of an i-motif are base pairs stacked face-to-face with each other.^[9] An i-motif unfolds and dissociates in an alkaline environment but refolds if the environment changes back to acidic.^[10] Therefore, the i-motif can be used as a molecular machine, switching between unfolded random strings and folded i-motif structures.^[11–13] Understanding the reversible folding and unfolding mechanisms of the i-motif is essential for its applications.

Due to the lack of experimental data on cytosine-base interactions, the nature of i-motif formation and dissociation mechanisms remains obscure. Gas-phase experiments can reliably reveal complex characteristics,^[14,15] but these experiments are very difficult even for H-bonded DNA base pairs. Observation of protonated cytosine–cytosine ($C-C^+$) base pairs outside the i-motif is challenging, so only gas-phase data exist for the base pairs of cytosine derivatives.^[16] Other methods, such as base association in polar and nonpolar solvents, on electrodes and with X-rays, provide only indirect in-

formation about the cytosine-base interactions.^[17]

In this work, we use *ab initio* quantum chemical calculations^[18] to investigate the formation and dissociation mechanisms of a $C-C^+$ base pair. Our calculations suggest that, in an acidic environment, a cytosine monomer is first protonated and then dimerized with an unprotonated cytosine monomer to form a $C-C^+$ base pair; in an alkaline environment, a $C-C^+$ base pair is first unprotonated and then dissociated into two cytosine monomers. Our *ab initio* calculations on $C-C^+$ base pairs provide useful information for understanding the molecular origin of the formation and dissociation of i-motif; our results can be systematically refined when much larger computational resources are available.

2. Computational method

To obtain energy differences between configurations, nine molecular structures were set up and run through *ab initio* calculations: (i) two cytosine monomers; (ii) a $C-C^+$ base pair; (iii) two protonated cytosine monomers; (iv) an unprotonated cytosine–cytosine dimer; (v) a water molecule; (vi) a hydronium ion; (vii) a hydroxyl ion; (viii) a $C-C^+$ base pair with the hydrogen ion located at different places between the two 3' end nitrogen atoms; and (ix), a hemiprotonated cytosine dimer with different distances between the two 3' end nitrogen atoms. The cytosine monomer and dimer configurations were taken from the NMR structure of the i-motif (PDB ID:1YBN).^[19] The initially constructed molecular structures were optimized by the *ab initio* calculations at the HF/6-

*Project supported by the National Basic Research Program of China (973 Program, Grant No. 2013CB932804), the National Natural Science Foundation of China (Grant Nos. 91227115 and 11121403), and the Hundred Talent Program of the Chinese Academy of Sciences (CAS).

†Corresponding author. E-mail: wangyt@itp.ac.cn

311g** level; the total energies of the optimized structures were calculated at both the HF/6-311g** and MP2/6-311g** levels, plus another optimization and energy calculation procedure at the B3LYP/6-311g** level was made for comparison. All *ab initio* calculations were performed using the Gaussian software.^[18]

3. Results and discussion

3.1. Optimized C–C⁺ base pair configuration

According to its NMR structure shown in Fig. 1(a), i-motif is constructed by the stacking of C–C⁺ base pairs. However, since the hydrogen atoms cannot be identified by the NMR scanning, the exact location of the proton in the C–C⁺ base pair is unknown. Therefore, we added a proton between the 3' end nitrogen atoms; the optimized dimer structure is shown in Fig. 1(b). In the optimized structure, the distance between the two 3' end nitrogen atoms is 2.89 Å

the hydrogen ion, lined up with the two nitrogen atoms, is 1.02 Å away from one 3' end nitrogen atom and 1.87 Å away from the other one. The energy landscape obtained by moving the proton along the line between the two nitrogen atoms, as shown in Fig. 2, is a double well with an energy barrier of about 20 kJ/mol calculated at the HF and B3LYP levels (about 45 kJ/mol at the MP2 level). These results indicate that the hydrogen atom prefers to stay close and chemically bonded to either of the two 3' end nitrogen atoms, which is consistent with the NMR results.^[20]

3.2. Formation and dissociation of C–C⁺ base pair

The formation and dissociation of C–C⁺ base pairs are crucial for understanding the folding and unfolding mechanisms of the i-motif, as C–C⁺ base pairs are the building blocks of an i-motif. In this work, we propose a reversible chemical-reaction mechanism for the formation and dissociation of a C–C⁺ base pair based on our *ab initio* calculations. The postulated chemical reaction paths are shown in Fig. 3.

Two possible C–C⁺ base-pair formation paths in an acidic environment are illustrated in Fig. 3(a). Since hydronium ions are abundant in an acidic environment and easily lose hydrogen atoms, it is natural to regard the formation of C–C⁺ base pairs from two unprotonated cytosine monomers and one hydronium ion. In the first proposed reaction path, the two cytosine monomers first attract each other to form an unprotonated cytosine dimer and then one hydrogen atom is detached from a hydronium ion and attached to one of the 3' end nitrogen atoms; in the second reaction path, one hydrogen atom is detached from a hydronium ion and attached to one of the cytosine monomers before the protonated cytosine and the unprotonated cytosine attract each other to form a C–C⁺ base pair. The configurations with different combinations of those molecules went through *ab initio* calculations; the obtained reaction enthalpies for various steps are listed in Table 1. As shown in Table 1, the total reaction enthalpy for a cytosine monomer and a hydronium to form a C–C⁺ dimer and a water molecule, calculated by the MP2 method, is about –747.121 kJ/mol (–742.696 kJ/mol by HF and –743.172 kJ/mol by B3LYP), which suggests that, in an acidic environment, it is energetically favorable for two cytosine monomers to snatch a proton from hydronium and form a C–C⁺ base pair. Moreover, the second reaction path mentioned above is energetically more favorable than the first one, since it costs energy for two cytosine monomers to form an unprotonated cytosine dimer (ΔH_{11}^a), while it releases energy for a cytosine monomer to snatch a proton from hydronium (ΔH_{21}^a) before another energy-releasing process takes place to form the C–C⁺ base pair. This is consistent with the experimental results,^[21] showing that i-motifs only fold in an acidic environment but unfold even in a weak alkaline environment.

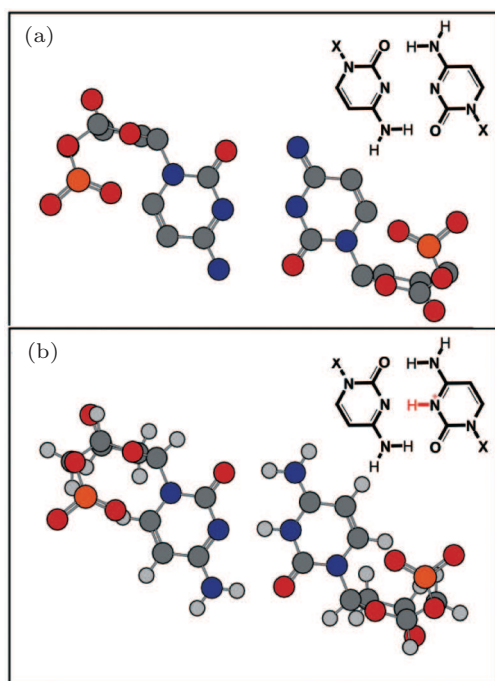


Fig. 1. (color online) (a) Cytosine dimer configuration taken from the i-motif NMR structure without the central hydrogen atom. (b) Optimized hemiprotonated cytosine dimer configuration with the central hydrogen atom.

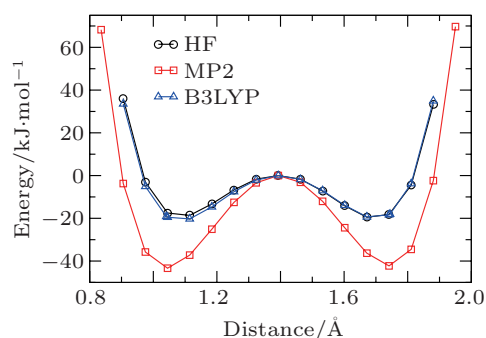


Fig. 2. (color online) Energy landscape of the hemiprotonated cytosine dimer with the hydrogen ion moving along the line between the two 3' end nitrogen atoms.

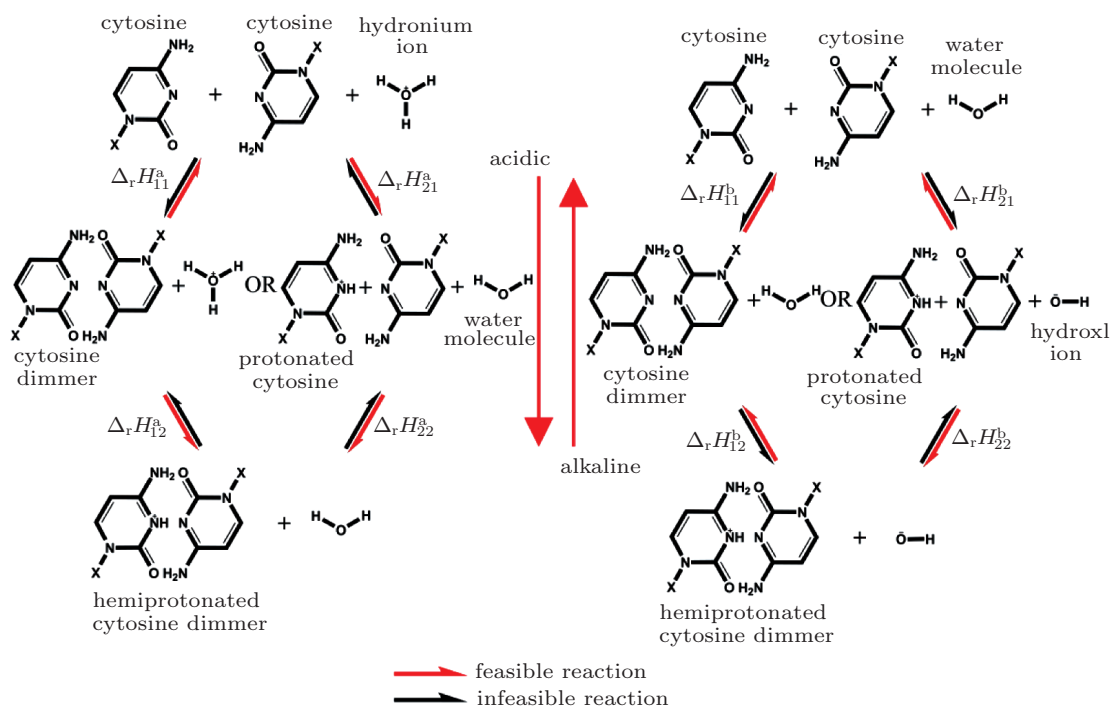


Fig. 3. (color online) (a) Two possible chemical reaction paths for the formation of a C-C⁺ base pair. The reactants are on the top. (b) Two possible chemical reaction paths for a C-C⁺ base pair to dissociate. The reactants are at the bottom.

Table 1. Calculated reaction enthalpies for the reaction paths in Fig. 3.

Reaction enthalpy	HF/(kJ/mol)	MP2/(kJ/mol)	B3LYP/(kJ/mol)
ΔH_{11}^a	82.229	58.512	85.354
ΔH_{12}^a	-824.925	-805.634	-828.526
ΔH_{21}^a	-546.274	-524.526	-538.595
ΔH_{22}^a	-196.421	-222.596	-204.577
ΔH^a	-742.696	-747.121	-743.172
ΔH_{11}^b	82.229	58.512	85.354
ΔH_{12}^b	231.670	266.593	230.715
ΔH_{21}^b	510.321	547.701	520.646
ΔH_{22}^b	-196.421	-222.596	-204.577
ΔH^b	313.899	325.106	316.069

In contrast, two possible C-C⁺ base-pair dissociation paths in an alkaline environment are illustrated in Fig. 3(b). Hydroxyls are considered to participate in the dissociation reaction since they are abundant in an alkaline environment. In the first proposed reaction path, the C-C⁺ base pair first gives a proton to the hydroxyl before the unprotonated cytosine dimer dissociates into two unprotonated cytosine monomers; in the second reaction path, the C-C⁺ base pair first dissociates into one protonated cytosine monomer and one unprotonated cytosine monomer and then the protonated cytosine monomer gives out a proton to the hydroxyl. Our calculations indicate that the total reaction enthalpy for a C-C⁺ base pair to give up a proton to the hydroxyl and dissociate into two cytosine monomers is negative (-325.106 kJ/mol with MP2, -313.899 kJ/mol with HF and -316.069 kJ/mol with B3LYP), suggesting that it is energetically favorable for a C-C⁺ base pair to dissociate in an alkaline environment. Accord-

ing to the data listed in Table 1, in the alkaline environment, a cytosine protonated cytosine dimer dissociates by giving up a proton to form an unprotonated cytosine dimer before dissociating into two cytosine monomers, which is the first proposed reaction path mentioned above. This result is consistent with the existing experimental results.^[21] From the energetic point of view, the energy released from the combination of proton and hydroxyl is larger than the energy for a C-C⁺ base pair to give up a proton; however, the energy released from the combination of a proton and a water molecule is much smaller. This ensures that the dissociation process can take place only in an alkaline environment.

4. Dissociation energy and force for detaching a C-C⁺ base pair in a vacuum

Since the i-motif can be used as a molecular machine reversibly switching between unfolded random strings and the folded i-motif structures,^[11-13] the energy and force needed to detach a C-C⁺ base pair is an essential topic of interest. Therefore, we varied the distance between the two 3' end nitrogen atoms and the energy and force were calculated for each distance. The consequent energy and force landscapes are shown in Fig. 4. As the distance increases, the energy monotonically increases but the increasing speed slows down (Fig. 4(a)). Correspondingly, the force decreases drastically with the distance (Fig. 4(b)) and an inversely proportional relationship was found between the force and the distance (Fig. 4(c)).

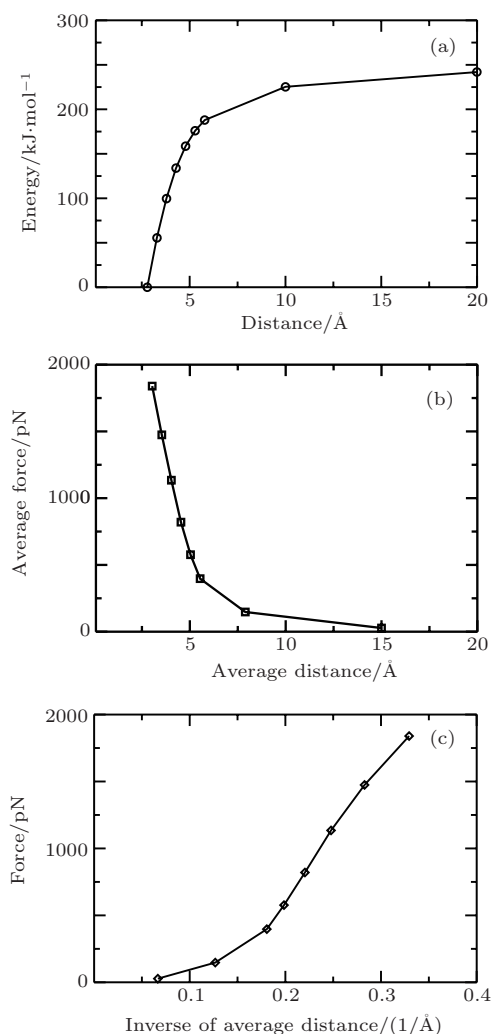


Fig. 4. (a) Energy landscape for detaching a C–C⁺ base pair along the line between the two 3' end nitrogen atoms. (b) The force landscape corresponding to panel (a). (c) Same plot as panel (b), but with the X-axis representing the inverse of the distance.

5. Conclusions

In summary, we have performed *ab initio* quantum chemical calculations to investigate the C–C⁺ base pair formation and dissociation mechanisms in acidic and alkaline environments, respectively, which is essential for understanding the transformation of an i-motif under different chemical conditions. Our calculations suggest that the hydrogen ion plays an important role in the formation and dissociation of C–C⁺ base pair. In an acidic environment, two cytosine monomers cannot form a dimer without a proton. One cytosine monomer has to be protonated by snatching a proton from a hydronium before it can combine with the other cytosine to form a C–C⁺ base pair. In contrast, in an alkaline environment, a C–C⁺ base pair gives up a proton to a hydroxyl and becomes an unprotonated cytosine dimer before it dissociates into two cytosine monomers. Our calculations also demonstrate that the force for detaching two cytosine monomers in the C–C⁺ base pair in a vacuum is inversely proportional to the distance between them.

It should be noted that our calculations assume greatly simplified conditions compared with experiments. Interactions between base pairs may further stabilize the i-motif but the environment, the configurational flexibility and thermal fluctuations may greatly reduce the reaction enthalpies. Nevertheless, the reaction paths we have determined and the mechanisms we have proposed provide an important qualitative understanding of the formation and dissociation of the i-motif and form the basis for future, more sophisticated theoretical studies of i-motif folding and unfolding mechanisms.

Acknowledgment

The authors thank the Supercomputing Center in the Computer Network Information Center at CAS for allocations of computer time.

References

- [1] Leroy J L, Guéron M, Mergny J L and Hélène C 1994 *Nucleic Acids Res.* **22** 1600
- [2] Leroy J L and Guéron M 1995 *Structure* **3** 101
- [3] Gehring K, Leroy J L and Guéron M 1993 *Nature* **363** 561
- [4] Chen L, Cai L, Zhang X and Rich A 1994 *Biochemistry* **33** 13540
- [5] Berger I, Kang C, Fredian A, Ratliff R, Moyzis R and Rich A 1995 *Nat. Struct. Mol. Biol.* **2** 416
- [6] Kang C H, Berger I, Lockshin C, Ratliff R, Moyzis R and Rich A 1994 *Proc. Natl. Acad. Sci. USA* **91** 11636
- [7] Kang C H, Berger I, Lockshin C, Ratliff R, Moyzis R and Rich A 1995 *Proc. Natl. Acad. Sci. USA* **92** 3874
- [8] Weil J, Min T P, Yang C, Wang S R, Sutherland C, Sinha N and Kang C H 1999 *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **55** 422
- [9] Spöner J, Leszczynski J, Vetterl V and Hobza P 1996 *J. Biomol. Struct. Dyn.* **13** 695
- [10] Jin K S, Shin S R, Ahn B, Rho Y, Kim S J and Ree M 2009 *J. Phys. Chem. B* **113** 1852
- [11] Liu D and Balasubramanian S 2003 *Angew. Chem. Int. Ed.* **42** 5734
- [12] Liu D, Bruckbauer A, Abell C, Balasubramanian S, Kang D J, Klenerman D and Zhou D 2006 *J. Am. Chem. Soc.* **128** 2067
- [13] Liu H, Xu Y, Li F, Yang Y, Wang W, Song Y and Liu D 2007 *Angew. Chem. Int. Ed.* **46** 2515
- [14] Yanson I K, Teplitsky A B and Sukhodub L F 1979 *Biopolymers* **18** 1149
- [15] Dey M, Moritz F, Grotemeyer J and Schlag E W 1994 *J. Am. Chem. Soc.* **116** 9211
- [16] Moehlig A R, Djernes K E, Krishnan V M and Hooley R J 2012 *Org. Lett.* **14** 2560
- [17] Choi J, Kim S, Tachikawa T, Fujitsuka M and Majima T 2011 *J. Am. Chem. Soc.* **133** 16146
- [18] Frisch M J, Trucks G W, Schlegel H B, Scuseria G E, Robb M A, Cheeseman J R, Montgomery J A, Vreven T, Kudin K N, Burant J C, Millam J M, Iyengar S S, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson G A, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox J E, Hratchian H P, Cross J B, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann R E, Yazyev O, Austin A J, Cammi R, Pomelli C, Ochterski J W, Ayala P Y, Morokuma K, Voth G A, Salvador P, Dannenberg J J, Zakrzewski V G, Dapprich S, Daniels A D, Strain M C, Farkas O, Malick D K, Rabuck A D, Raghavachari K, Foresman J B, Ortiz J V, Cui Q, Baboul A G, Clifford S, Cioslowski J, Stefanov B B, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin R L, Fox D J, Keith T, Laham A, Peng C Y, Nanayakkara A, Challacombe M, Gill P M W, Johnson B, Chen W, Wong M W, Gonzalez C and Pople J A 2004 *Gaussian 03, Revision C.02* (Wallingford CT: Gaussian, Inc.)
- [19] Esmaili N and Leroy J L 2005 *Nucleic Acids Res.* **33** 213
- [20] Lieblein A L, Kramer M, Dreuw A, Furtig B and Schwalbe H 2012 *Angew. Chem. Int. Ed.* **51** 4067
- [21] Kaushik M, Suehl N and Marky L A 2007 *Biophys. Chem.* **126** 154